

La recherche en évaluation : propos synthétiques et prospectifs

Jean-Marie De Ketele

Volume 29, numéro 1, 2006

L'actualité de la recherche en évaluation

URI : <https://id.erudit.org/iderudit/1086970ar>

DOI : <https://doi.org/10.7202/1086970ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

De Ketele, J.-M. (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et évaluation en éducation*, 29(1), 99–118.
<https://doi.org/10.7202/1086970ar>

Résumé de l'article

Cet article tente une synthèse de l'état de la recherche dans le domaine de l'évaluation, à partir des différentes contributions de ce numéro spécial et, plus largement, d'articles parus dans la revue *Mesure et évaluation en éducation* ou de communications présentées dans les colloques de l'ADMEE. Par suite de la multiplicité des travaux, cette synthèse tente de mettre de l'ordre en distinguant, d'une part, les pratiques évaluatives comme objets de la recherche et, d'autre part, l'évaluation comme outil de recherche. Pour chacun de ces aspects sont présentés les domaines qui ont été l'objet des plus grandes avancées et ceux qui soulèvent des problèmes insuffisamment abordés ou mal résolus.

La recherche en évaluation : propos synthétiques et prospectifs

Jean-Marie De Ketele

UCL, Louvain-la-Neuve

MOTS CLÉS : Évaluation, recherche, état de la question

Cet article tente une synthèse de l'état de la recherche dans le domaine de l'évaluation, à partir des différentes contributions de ce numéro spécial et, plus largement, d'articles parus dans la revue Mesure et évaluation en éducation ou de communications présentées dans les colloques de l'ADMEE. Par suite de la multiplicité des travaux, cette synthèse tente de mettre de l'ordre en distinguant, d'une part, les pratiques évaluatives comme objets de la recherche et, d'autre part, l'évaluation comme outil de recherche. Pour chacun de ces aspects sont présentés les domaines qui ont été l'objet des plus grandes avancées et ceux qui soulèvent des problèmes insuffisamment abordés ou mal résolus.

KEY WORDS : Evaluation, assessment, research, state of art

The aim of this article is to depict the state of the art in evaluation and assessment in education, based on three main sources that came out from this researchers community: (1) articles from this specific issue; (2) articles edited in the Journal Mesure et évaluation en éducation and (3) papers presented at international conferences of the Association pour le développement des méthodologies de l'évaluation en éducation (ADMEE). This leads us to distinguish between two main topics, for which we pointed out both main advancements, pitfalls, and questions. The first one considers evaluative practices as research objects and the second one, looks at evaluation as a research method.

PALAVRAS-CHAVE : Avaliação, investigação, estado da arte

Este artigo faz uma síntese do estado da investigação no domínio da avaliação, a partir das diferentes contribuições deste número especial e, ainda, de artigos publicados na revista Mesure et évaluation en éducation e de comunicações apresentadas nos colóquios da ADMEE. Na continuidade da multiplicidade dos trabalhos, esta síntese tenta pôr ordem distinguindo, por um lado, as práticas avaliativas como objectos de investigação e, por outro lado, a avaliação como utensílio de investigação. Para cada um destes aspectos são apresentados os domínios que foram objecto de maiores avanços e aqueles que levantam ainda problemas, por terem sido pouco explorados ou por estarem mal resolvidos.

Les propos tenus dans ce numéro spécial sur la recherche en évaluation sont intéressants à plus d'un titre. Cela tient notamment à la richesse des points d'entrée adoptés. Mais cela peut aussi donner l'impression d'un *patchwork*. Dans cette conclusion, nous allons essayer de démêler l'écheveau et de remettre les fils dans un certain ordre. Partir du thème de ce numéro spécial nous semble un bon principe ordonnateur : il s'agit de «recherche» et de «recherche en évaluation». Quelques rappels clarificateurs méritent d'être faits.

Il est bon de rappeler que si le chercheur est et doit être un penseur, il ne suffit pas d'être penseur pour être un chercheur ; on est éventuellement un excellent «essayiste», voire un «gourou» aux idées intéressantes. La recherche (Campbell & Stanley, 1966 ; De Bruyne, Herman & De Schoutheete, 1974 ; De Ketele & Roegier, 1991, 1993 ; Grawitz, 1976) suppose qu'un problème pertinent étant bien posé, un dispositif adéquat de recueil de données empiriques est conçu, une base de données est constituée dont on peut vérifier le degré de validité et de fiabilité, une méthodologie d'exploitation de celle-ci est précisée (elle doit être reproductible) et appliquée rigoureusement, des conclusions en résultent, une discussion de celles-ci tente de leur donner du sens. Ces caractéristiques définissent la recherche scientifique qui est loin d'être un processus linéaire et où le travail de penseur intervient surtout lors de la problématisation et de la re-problématisation (discussion et mise en perspective des conclusions).

Quand on parle de «recherche en évaluation», on pense immédiatement au champ de la recherche qui prend comme objet l'évaluation. Encore faut-il s'entendre sur l'objet. De la lecture des articles de ce numéro, on peut le cerner comme suit : il s'agit d'une pratique sociale, finalisée, située dans un contexte, impliquant une démarche de confrontation d'un référé à un référentiel (Figari & Achouche, 2001). Un premier travail de clarification consiste à dégager les différents types de pratiques évaluatives, susceptibles d'être pris comme objets de la recherche en évaluation, à partir de trois paramètres essentiels : les contextes, les fonctions et les démarches de l'évaluation. Mais il me semble aussi que parler de recherche en évaluation, c'est aussi ne pas oublier que l'évaluation est un outil méthodologique essentiel dans la recherche en éducation. En conséquence, notre tentative de mise en ordre abordera successivement la recherche sur l'évaluation, visant à améliorer notre connaissance

Note de l'auteur : Toute correspondance peut être adressée par courriel à l'adresse suivante : [jean-marie.deketele@psp.ucl.ac.be].

des pratiques évaluatives, et la recherche méthodologique visant à améliorer l'évaluation comme outil au service de la recherche en éducation. Ces deux préoccupations ne sont évidemment pas strictement indépendantes. Ainsi, par exemple, les enquêtes PISA sont des pratiques évaluatives qui se situent à un niveau macro, mais les épreuves PISA peuvent devenir des outils utilisés dans des recherches en éducation, comme celles qui visent à mettre à l'épreuve l'hypothèse selon laquelle les performances d'un système éducatif seraient différentes selon le degré de différenciation précoce de leurs curriculums¹. Dans cet exemple, il ne s'agit pas d'une recherche sur l'évaluation, mais d'une recherche en éducation prenant appui sur des données évaluatives.

Les pratiques évaluatives comme objets de la recherche en éducation

Les auteurs de ce numéro spécial constatent l'élargissement du champ de l'évaluation au fil du temps. Cet élargissement (voir aussi Figari & Achouche, 2001; Jorro, 2000; Vial, 2001) implique non seulement une multiplication des objets possibles de la recherche sur l'évaluation, mais entraîne une différenciation des problématiques, des méthodes de recueil et de traitement de l'information et de mise en perspective des conclusions. Négliger ce travail de différenciation, c'est entraîner la confusion et des inférences abusives.

Vers une typologie des objets

Si de nombreux paramètres sont susceptibles d'être utilisés pour différencier les pratiques évaluatives comme objets de recherche, trois paramètres semblent nécessaires pour caractériser l'évaluation, quelle qu'elle soit : le niveau du contexte, la fonction principale visée et le type de démarche utilisée. On pourrait s'étonner qu'un paramètre comme celui des acteurs ne soit pas mentionné : il est certes important, mais il est niché dans les paramètres précités.

Les pratiques évaluatives sont des pratiques sociales situées². Leur étude implique donc que l'on distingue le niveau de contexte où elles se situent : micro, méso et macro. Le niveau micro implique que l'on se situe au sein d'un processus d'enseignement-apprentissage ou de formation (le niveau de la classe ou du groupe de formation). Le niveau méso implique que l'on se situe au sein d'un établissement ou d'une institution d'enseignement ou de formation. Le niveau macro implique que l'on se situe sur le plan d'un système éducatif ou de formation.

Les pratiques évaluatives sont des actions finalisées, c'est-à-dire qu'elles sont censées servir une fonction prioritaire. Au fil du temps, de nombreux concepts ont été utilisés pour nommer ces fonctions et il importe de souligner que certaines distinctions conceptuelles ont été source de confusion. Ainsi, par exemple, la distinction entre l'évaluation formative et sommative due à Scriven est malheureuse, car elle confond ce qui est de l'ordre de la fonction (le «pour quoi?») et ce qui est de l'ordre de la démarche (le «comment?»). Ainsi, l'évaluation formative nomme la fonction d'amélioration de l'action d'apprentissage en cours, tandis que l'évaluation sommative nomme la démarche qui consiste à faire une somme d'éléments (attribution d'une note). La distinction de Bloom, qui consiste à compléter la distinction de Scriven par l'évaluation diagnostique pour parler de l'évaluation permettant de faire un diagnostic pour préparer un nouvel apprentissage, est aussi source de confusion, car le diagnostic est aussi un processus fondamental au sein de l'évaluation formative. Bloom percevait cependant, à travers sa triple distinction malheureuse, trois fonctions principales différentes de l'évaluation : préparer adéquatement une nouvelle action ; améliorer une action en cours ; certifier au terme d'une action. Il nous paraît pertinent d'utiliser les concepts suivants pour nommer sans ambiguïté ces trois fonctions prioritaires, à savoir respectivement : l'évaluation d'orientation, l'évaluation régulatrice et l'évaluation certificative. Il est à noter que le concept d'évaluation régulatrice est moins restrictif que celui d'évaluation formative, car ce dernier se restreint au contexte micro ; par ailleurs, certains auteurs tiennent à dire que l'évaluation régulatrice au sein d'un apprentissage ou d'une formation peut être formative ou formatrice (les chercheurs de l'École d'Aix-en-Provence ont beaucoup exploité cette distinction).

Actions situées et finalisées, les pratiques évaluatives se caractérisent par trois types de démarches. La plus connue, souvent la seule pratiquée consciemment par de nombreux enseignants, c'est la démarche sommative. Elle consiste à travers des sommations³ à évaluer le niveau de performance. Par un chiffre ou une appréciation hiérarchique (comme une mention par exemple), on essaie de traduire un bilan par rapport à une norme soit de groupe (ce qui est le plus fréquent malheureusement), soit de maîtrise d'un objectif, soit de progrès par rapport à un niveau initial. Une même note ou mention obtenue par deux personnes différentes semble signifier qu'elles sont équivalentes pour ce qui est de la performance. On sait qu'il n'en est rien. Dans une épreuve de vingt questions, elles peuvent toutes deux obtenir une note de 10 sur 20 en réussissant chacune dix questions différentes. Contrairement à la démarche sommative, la démarche descriptive se centre sur la distinction et la description

précise de ce qui est maîtrisé ou non, des difficultés observées, des procédures heureuses ou malheureuses utilisées par l'apprenant. Une troisième démarche complète les deux démarches évoquées : la démarche herméneutique, appelée souvent aussi intuitive (terme perçu comme moins «pédant» aux yeux de certains acteurs de terrain). Cette démarche consiste à donner du sens à une série d'indices recueillis de façon plus ou moins consciente. Elle comprend donc une part importante de «subjectivité». Comme le signale De Ketele dans *L'évaluation conjuguée en paradigmes* (1993), elle a donc été vilipendée pendant la période dominée par le paradigme docimologique, mais elle a été réhabilitée par Jacques Weiss dans un article célèbre *La subjectivité blanchie* (1984). Un phénomène de réductionnisme historique a conduit à restreindre la démarche sommative à la fonction certificative, la démarche descriptive à la fonction régulatrice ou formative, la démarche herméneutique ou intuitive à la fonction d'orientation. En fait, les trois démarches peuvent être utilisées, de façon pertinente ou non, quelle que soit la fonction. On peut très bien certifier un apprentissage par un bulletin descriptif ou une formation professionnelle par un certificat décrivant les compétences apprises et validées. Une interrogation notée peut donner des indications utiles pour améliorer l'apprentissage en cours ; de même, l'intuition du formateur sur le sens à donner à un événement vécu peut se révéler très utile pour débloquer une formation en cours. Passer un test ou faire distinguer et décrire des tâches où des difficultés apparaissent, ce sont des démarches qui peuvent permettre de prendre des décisions d'orientation pertinentes. Notons en fait que pour une même fonction, il peut être utile et pertinent de combiner plusieurs démarches.

Le tableau suivant croise les trois paramètres (les trois types de contextes, les trois fonctions et les trois types de démarches) et permet de visualiser les différents types d'objets susceptibles d'être pris en considération par la recherche sur les pratiques évaluatives.

Tableau 1
Les objets possibles de la recherche sur l'évaluation

<i>CONTEXTES</i>	<i>DÉMARCHES</i>	<i>FONCTIONS</i>		
		<i>ORIENTATION</i>	<i>RÉGULATION</i>	<i>CERTIFICATION</i>
MICRO	Sommative	Mi S O	Mi S R	Mi S C
	Descriptive	Mi D O	Mi D R	Mi D C
	Herméneutique	Mi H O	Mi H R	Mi H C
MÉSO	Sommative	Mé S O	Mé S R	Mé S C
	Descriptive	Mé D O	Mé D R	Mé D C
	Herméneutique	Mé H O	Mé H R	Mé H C
MACRO	Sommative	Ma S O	Ma S R	Ma S C
	Descriptive	Ma D O	Ma D R	Ma D C
	Herméneutique	Ma H O	Ma H R	Ma H C

Quelques questions prioritaires pour la recherche

Les objets des recherches sur l'évaluation sont nombreux et diversifiés. Au terme de ce numéro spécial, il nous paraît important d'épingler quelques-unes des questions prioritaires à traiter ou à approfondir dans les recherches à venir. Celles-ci balayent toutes les plages de notre tableau, même si certains secteurs ont concentré l'attention des chercheurs.

Si l'on se situe dans les contextes micro, il faut reconnaître que peu de chercheurs ont fait porter leurs recherches sur les problèmes posés par la fonction d'orientation, ce qui est sans doute aussi un signe du peu d'attention qui lui est accordée par les acteurs de terrain⁴. Parmi les questions prioritaires qui se posent, nous pointerons les suivantes :

- Quel est l'impact de l'absence ou, au contraire, de la présence d'attention à l'évaluation des préalables au début d'une nouvelle séquence d'apprentissage, et tout particulièrement au début d'une nouvelle année scolaire ? Nous avons dirigé plusieurs recherches qui montrent qu'en moyenne 25% des acquis évalués au terme de la scolarité primaire ne sont plus maîtrisés par ces mêmes élèves deux mois et demi après, c'est-à-dire à l'entrée du collège (Cnudde & De Ketele, 1986; Cnudde, De Ketele, Delande & Delory, 1988; Delannay & De Ketele, 2004; Delory, 1993). Autour de cette question centrale se greffent de nombreuses questions, comme celle relative au type de démarche qui se montre la plus efficiente ou celle relative à l'impact sur la réduction ou l'aggravation des inégalités scolaires.

- Dans quelle mesure les pratiques évaluatives à fonction d'orientation tiennent-elles suffisamment compte, non seulement des besoins des apprenants ou des personnes en formation, mais des potentialités de chacun d'entre eux ? Chargé de faire la synthèse des travaux lors du vingtième anniversaire des ZEP (Zones d'éducation prioritaire) et des REP (Réseaux d'éducation prioritaire) en France, j'ai pris conscience de l'importance de ce type de question. En effet, parmi les déterminants qui font la différence entre les enseignants ou les équipes d'enseignants qui font ou, au contraire, ne font pas faire des progrès significatifs à leurs élèves, c'est le fait que les enseignants parviennent à faire de chacun des élèves un « pôle d'excellence sur quelque chose » (De Ketele, 2002). Déceler les potentialités différenciées de chaque élève semble d'autant plus fondamental que l'élève cumule les handicaps.

Dans les contextes micro, les chercheurs des deux dernières décades se sont penchés à maintes reprises sur la fonction régulatrice de l'évaluation, au point de voir les distinctions se multiplier : évaluation formative et formatrice (Nunziati, 1990 ; Scallon, 1988) ; évaluation rétroactive, interactive et proactive (Allal, 1979) ; autoévaluation, hétéroévaluation et coévaluation (Paquay, 2004) ; autoévaluation entre autocontrôle et autoquestionnement (Vial, 1996, 2001), etc.

Toutefois, des questions prioritaires restent sans réponse suffisante :

- Comment se fait-il que, malgré l'abondance de travaux sur la fonction régulatrice de l'évaluation, celle-ci semble pénétrer très peu les pratiques des enseignants au sein de leur classe ? Si les recherches se sont beaucoup concentrées sur l'évaluation des produits, des procédures et des processus utilisés par les élèves en cours d'apprentissage ou les personnes en formation, et s'il est vrai qu'il reste un énorme travail à faire dans ce domaine en liaison plus étroite avec les objets enseignés, ne devraient-elles pas se centrer aussi davantage sur l'exploitation (et le manque d'exploitation) de l'information recueillie par les enseignants ou les formateurs ?
- L'évaluation qui vise à améliorer l'apprentissage en cours, c'est-à-dire régulatrice, ne devrait-elle pas se préoccuper fondamentalement de la façon dont est géré le temps (temps effectif d'enseignement et d'apprentissage, temps consacré à la gestion de la discipline, temps perdu) et de la façon dont les diverses activités sont rythmées plus ou moins optimalement ?

Comme le rappelle Dauvisis dans cet ouvrage et ailleurs (2001), la docimologie a multiplié, entre 1950 et 1980, les recherches sur les examens et les déterminants de l'élaboration de la note, c'est-à-dire sur la démarche sommative dans le cadre de la fonction certificative. Cette approche docimologique a été complétée par des approches plus psychologiques, comme celle de Noizet et Caverni (1978) et plus récemment de Bressoux et Pansu (2003). Tout en étant nécessaires, ces approches souffrent de deux «négligences» fondamentales. En se centrant trop exclusivement sur la fiabilité (note identique selon les personnes et les circonstances), elle oublie de se poser la question bien plus fondamentale de la pertinence de ce qu'il faut évaluer (et apprendre) pour certifier à bon escient. Une seconde négligence réside dans le fait, non seulement d'assimiler la certification à la démarche sommative, mais à l'assimiler exclusivement à une démarche de classement par rapport à une norme de groupe (voir les critiques de Cardinet à ce propos) et donc aux présupposés de la courbe normale des résultats (voir aussi la critique de Bloom). Dans cet esprit, nous voudrions mettre en évidence deux questions qui nous paraissent prioritaires pour les travaux ultérieurs :

- S'il est vrai qu'il importe de certifier des compétences⁵, plutôt que de simples savoirs et savoir-faire (applications), la démarche sommative peut-elle traduire en une note (donc sur une échelle mathématique unidimensionnelle par définition) la maîtrise d'une compétence qui, par définition, est multidimensionnelle (De Ketele & Gérard, 2004 et à paraître)? Fondé sur les postulats de l'unidimensionnalité et de la distribution normale, la théorie des réponses aux items peut-elle continuer à valider les épreuves d'évaluation certificative, dans la mesure où elles ont la prétention d'évaluer des compétences?
- Ne devrait-on pas se pencher davantage sur les démarches descriptives et herméneutiques pour certifier des compétences ou des tâches complexes? Comme le souligne Vial dans ce numéro, quelle est la part du processus et du produit à prendre en considération dans la certification des compétences? Ne faut-il pas créer une conceptualisation et une méthodologie de la certification qui s'appuie sur les démarches descriptive et herméneutique?

Les contextes méso, les établissements scolaires et les institutions de formation, ont été surtout préoccupés par l'élaboration des projets d'établissement. Il ne faut cependant pas oublier deux grands débats. Le premier a trait au problème de l'orientation: faut-il orienter de façon précoce les élèves ou faut-il un tronc commun qui couvre intégralement l'école et le collège (à

savoir jusqu'à 15 ans)? Le second a trait aux réformes des institutions et des programmes de formation des enseignants et même, dans certains pays, à l'allongement de la formation initiale. Ces deux grands débats concernent les pratiques évaluatives au sein des établissements d'enseignement et de formation pour orienter respectivement les élèves à travers les filières d'études et les programmes de formation des enseignants. À nos yeux, deux questions prioritaires devraient être abordées par les recherches sur l'évaluation :

- Quelles sont les pratiques évaluatives menées, tout particulièrement au sein des collèges, pour orienter les élèves dans leur parcours scolaire? Quel est l'impact différencié des différents types de pratiques d'orientation? En France, les travaux de différents organes, comme la Commission Thélot (Commission sur l'avenir de l'école en France) ou le Haut Conseil d'évaluation de l'école, arrivent à la même conclusion: les pratiques actuelles d'orientation des élèves sont non seulement inefficaces sur bien des points de vue, mais contribuent à aggraver les inégalités scolaires. Des travaux scientifiques récents, comparant plusieurs systèmes éducatifs, vont dans le même sens (Dupriez & Dumay, 2005; Hanushek & Wöbman, 2005).
- Alors que, ces dix dernières années, les travaux sur le développement professionnel et sur la professionnalisation des enseignants se sont multipliés, alors que les facultés de médecine et les écoles polytechniques ont instauré de nouveaux programmes de formation basés sur les approches par problèmes et par projets, comment se fait-il que les établissements de formation des enseignants continuent, dans leurs tentatives de réformes, à établir des programmes constitués d'une juxtaposition de disciplines diverses et de stages sans guère de liens avec les aspects théoriques de la formation? N'est-il pas temps de mener des recherches empiriques visant à déterminer, à partir d'une analyse des contextes de pratiques, des référentiels des métiers de l'enseignement et, à partir de là, des référentiels de compétences, des référentiels de formation initiale et continuée et des référentiels d'évaluation? Ou va-t-on continuer à développer chez les futurs enseignants des «savoirs ignares» (selon l'expression d'Edgard Morin), c'est-à-dire des savoirs non utilisés dans la pratique professionnelle, comme le montrent les enquêtes auprès des enseignants?

L'évaluation régulatrice au sein des établissements scolaires mérite l'attention des chercheurs sur deux aspects prioritaires à nos yeux, si l'on prête attention à certains travaux comme le colloque du vingtième anniversaire des ZEP et des REP (Zones et Réseaux d'éducation prioritaire) ou la vaste enquête

menée au sein des établissements scolaires par une équipe pluridisciplinaire de chercheurs belges sous la direction de Maroy (2002). Il s'agit de la nécessaire, mais combien insuffisante à l'heure actuelle, concertation entre les enseignants, ainsi que le travail en équipes verticales et horizontales. Il s'agit aussi, plus spécifiquement, des conseils, comme le conseil de classe, dans leur fonction régulatrice; il est étonnant de voir combien la recherche sur les pratiques évaluatives n'a pas prolongé les travaux menés dans les années 80. En relation avec ces deux aspects, les questions suivantes nous paraissent prioritaires:

- Ne faut-il pas une radioscopie des pratiques de concertation et de travail en équipe au sein des établissements, tant sur les aspects quantitatifs et qualitatifs? Qu'est-ce qui différencie les établissements à ce propos? En quoi ces pratiques sont-elles de véritables processus d'évaluation régulatrice aboutissant à des prises de décisions, elles-mêmes objet d'un suivi?
- Ne faut-il pas aussi une radioscopie des conseils de classe? Quels en sont les objectifs réels? les modalités effectives? les activités de préparation? les activités de suivi? Sont-ils restés des instances de jugement négatif (voir Frères & De Ketele, 1985; Marchand, 1979)? ou sont-ils devenus de réelles instances d'évaluation régulatrices et de prises de décision collectives?

L'évaluation certificative des établissements scolaires ou de formation a fait l'objet de peu de recherches, alors que se dessine une tendance de plus en plus forte pour le rendement de compte. En Angleterre, une instance indépendante évalue les établissements scolaires; elle est susceptible au besoin de prendre des mesures draconiennes, y compris la fermeture ou le remplacement de l'équipe enseignante, en cas de performances scolaires jugées insuffisantes de façon récurrente. Dans de nombreux pays, on assiste à un phénomène de «judiciarisation» au sein du système éducatif, à un point tel que des avocats se spécialisent en cette matière⁶. Par ailleurs, le système éducatif réclame toujours plus de moyens à l'État et aux collectivités régionales ou locales. On assiste donc en retour à une demande de rendement de compte sur la plus-value apportée par les compléments de ressources abordées. Ces évolutions nous conduisent à nous poser deux ordres de questions: la certification au sein de l'école et la certification des écoles elles-mêmes:

- Si l'évaluation certificative des performances des élèves par l'enseignant a fait l'objet de nombreux travaux, par contre l'évaluation certificative collégiale, en l'occurrence le jury, a été négligée par les chercheurs. La possibilité de recours offerte aux parents attire maintenant l'attention sur le fonctionnement des jurys et il est donc temps de mener des recherches

empiriques sur ce sujet. Quels sont les critères de prise de décision? Quels sont les jeux de pouvoir? Quel est le rôle respectif des performances scolaires et des comportements en classe ou au sein de l'école? Qui introduit un recours? Pourquoi? Avec quels effets?

- Quelles sont les significations possibles du rendement de compte pour un établissement scolaire ou de formation? Rendre compte sur quoi? Faut-il, comme la Commission Thélot le suggère en France, se fixer «un socle commun de connaissances, de compétences et de comportements» à maîtriser par 100% des élèves⁷ et à partir duquel tout enseignant et tout établissement doit rendre compte? Qui définit ce socle? Un socle pour construire quoi? Y aurait-il deux niveaux pour l'évaluation certificative: l'évaluation du socle que doivent maîtriser 100% des élèves et l'évaluation d'un au-delà du socle, qui permettrait de différencier les élèves et qui serait déterminant pour leur orientation?

Au fil du temps, la problématique de l'évaluation s'est progressivement élargie au niveau méso d'abord, au niveau macro ensuite. Le célèbre modèle CIPP (Contexte, Input, Processus, Produit) de Stufflebeam (1971) a été le moteur de cet élargissement. Actuellement, les États et les organismes internationaux se préoccupent de plus en plus de l'évaluation des systèmes éducatifs. Ici encore, les recherches devraient pouvoir porter sur les trois fonctions: orienter les systèmes éducatifs en fonction des enjeux; les réguler; les certifier. Orienter un système éducatif consiste d'abord à préciser les enjeux sociétaux d'un monde se complexifiant et à évaluer les ressources et les contraintes de l'environnement pour concevoir et implanter des curriculums pertinents (au sens le plus large du terme et non simplement réduit à des programmes d'enseignement). La fonction d'orientation au niveau macro mériterait davantage d'études:

- Comme le souligne Roegiers (1997), quelle logique sous-tend l'élaboration des curriculums? une logique de l'expertise? une logique stratégique? une logique de projet? Quel est le statut et le rôle des personnes qui interviennent dans cette élaboration? En quoi et dans quelle mesure l'évaluation du contexte intervient et fonde les nouveaux curriculums? Dans quelle mesure les évaluations internationales jouent-elles un rôle de pression?
- La plupart des pays ont mis ou sont en train de mettre des instances de pilotage de leur système éducatif, comme le souligne Demeuse dans ce numéro. Quels sont les fondements et les pratiques qui président à la conception des politiques de pilotage selon les pays? Quels indicateurs

sont privilégiés par les uns et par les autres? Quel sens est attribué au concept de qualité (voir les Actes du Colloque de l'ADMEE à Lausanne en 2003)? Les batteries d'indicateurs privilégient-elles l'efficacité ou l'efficience ou l'équité? La conception du pilotage permet-il «un contrôle citoyen» (selon l'expression de Demeuse)?

À quoi sert la mise en place d'une politique de pilotage si elle ne contribue pas à réguler effectivement le système éducatif? Deux aspects nous semblent prioritaires pour les recherches à venir :

- Quels sont les modes de communication utilisés par les responsables du pilotage? Quels acteurs touchent-ils? Avec quel degré de réceptivité? N'y a-t-il pas des instances qui court-circuitent certaines données? Les données récoltées et communiquées ne sont-elles pas lues et interprétées différemment selon les acteurs? Qu'est-ce qui est fait aux différents paliers du système éducatif pour dépasser le simple constat?
- Dans la plupart des pays, les «interfaces» (inspecteurs, conseillers pédagogiques, formateurs, coordinateurs, personnes ressources, etc.) se sont multipliés. Quels rôles jouent-ils dans la régulation, tant sur le plan de la prise d'information, que de la communication, de l'intervention et de l'accompagnement? Quelles sont leurs pratiques? Comment sont-ils reçus? sollicités? évalués?

On a vu, ces dernières années, se multiplier les enquêtes évaluatives internationales. Des organismes internationaux comme l'OCDE conduisent régulièrement des audits nationaux ou se livrent à des accréditations de programme. Ce sont là des évaluations certificatives au plan macro. Elles ne sont pas sans poser des questions :

- Comment est déterminé le dénominateur commun de ce qui doit être évalué dans des pays tellement différents sur le plan des programmes, des modes de fonctionnement et des fondements culturels? Les enquêtes internationales ne jouent-elles pas un rôle de «normalisation»? Ne doivent-elles pas être complétées par des dispositifs d'évaluation d'une part plus différenciés, d'autre part prenant en compte des objets non évaluables ou moins facilement évaluables par des épreuves écrites communes?
- Quels sont les fondements des différentes formes d'audits des systèmes éducatifs? Quelles en sont les méthodologies? Quelle est la part du regard externe et des regards internes? Des standards communs se développent-ils progressivement dans les instances d'évaluation et d'accréditation?

L'évaluation comme outil de la recherche

Si l'évaluation est objet de la recherche, elle est aussi un outil indispensable dans de nombreuses recherches sur l'éducation. Pour bien clarifier la distinction entre la recherche sur l'évaluation et l'évaluation comme outil de recherche, partons de la recherche menée par Jorro (2002) et dont les conclusions sont rapportées dans ce numéro. Partant de la diversité des pratiques évaluatives, celle-ci s'est posé la question de savoir si le positionnement des acteurs par rapport à l'évaluation et si les ruptures entre le dire et le faire ne permettraient pas de dresser une typologie des rapports de ceux-ci à l'évaluation. Pour apporter une réponse à cette question, elle a mis au point un dispositif de recueil d'information et d'analyse de contenu⁸. Cette méthode a permis de distinguer trois types de rapports à l'évaluation, appelés respectivement le rapport d'«improvisation», le rapport «industriel» et le rapport «clandestin». Il s'agit bien d'une recherche sur l'évaluation. Imaginons qu'elle poursuive sa recherche en se posant maintenant la question de savoir si le type de rapport à l'évaluation détermine, dans une certaine mesure, les niveaux de performances scolaires pour ce qui est des plus-values et de la réduction des inégalités. Pour répondre à cette question, le chercheur sera amené à construire et à valider un outil d'évaluation pour «mesurer» les plus-values et l'éventuelle réduction des inégalités. Ceci suppose, d'une part, une réflexion fondamentale sur quels objets évaluer et un travail méthodologique de construction et de validation de l'outil d'évaluation. Outre la recherche sur l'évaluation, la recherche en évaluation comprend aussi tout un travail sur le choix, la conception et la validation des outils d'évaluation utilisés dans la recherche en éducation. Passons en revue quelques lignes de recherches qui nous paraissent prioritaires pour les années à venir.

De nombreuses recherches en éducation ont pour ambition de tester l'influence de certaines «variables d'entrée» (comme le niveau socio-économique et le genre, par exemple) ou de «variables processus» (comme le type de rapport à l'évaluation de l'enseignant ou le degré d'engagement de l'étudiant, par exemple) sur des «variables effets» (comme la plus-value pour ce qui est des termes de performances scolaires en une période de temps ou la réduction des inégalités, par exemple). Dans la plupart de ces recherches, les variables effets sont «mesurées» par des épreuves d'évaluation. Comme déjà signalé, avant d'être un problème méthodologique, il s'agit d'un problème politique: le choix du type d'effets à évaluer est fondamental. S'agit-il de connaissance à restituer fidèlement? de la maîtrise des savoir-faire de base

(application de règles ou d'algorithmes)? de compétences (résolution de problèmes ou effectuation de tâches complexes)? de savoir-être (systèmes de valeurs ou attitudes ou comportements habituels intériorisés)? de savoir-devenir (ce qui est de l'ordre du projet)? La méthode de construction et de validation n'est pas la même selon les différents types d'effets pris en compte. Si les travaux éducatifs ont fait des avancées considérables sur la conception et la validation des épreuves écrites ayant trait à l'évaluation des savoirs et des savoir-faire de base, il n'en est rien sur l'évaluation des compétences (voir la critique déjà signalée plus haut). D'un point de vue éducatif, il n'existe guère de travaux sur l'évaluation du développement des savoir-être. Certes, les psychométriciens ont construit et validé un certain nombre de tests, mais ceux-ci portent sur les aspects structurels (c'est-à-dire foncièrement stables, et relevant donc de la personnalité profonde) de certains savoir-être. Mais dans le processus éducatif ou de formation, l'enseignant ou le formateur cherche à faire passer dans «l'habitude intériorisée» des comportements ou des savoir-faire souhaitables (par exemple, la précision, la prise de distance par rapport à ses pratiques, le souci constant de se poser les questions pertinentes, etc.). Trop souvent, les variables effets des recherches négligent de prendre en considération les savoir-être développés par l'apprentissage ou la formation. Sans doute, l'absence de travaux méthodologiques explique en partie cette lacune. C'est donc là un champ important à explorer.

Ces vingt dernières années, les recherches en éducation visent à évaluer l'efficacité de démarches d'apprentissage ou de formation (niveau micro) ou l'efficacité des établissements (niveau méso) ou l'efficacité des systèmes éducatifs (niveau macro). Des méthodes de plus en plus sophistiquées ont été mises au point. Les premières avancées sont dues à Mac Guigan (in d'Hainaut, 1975) et à d'Hainaut (1975) qui ont mis au point des indices de gain brut et relatif.

Ces méthodes exigent une double comparaison: la comparaison des mesures des performances finales à une ou plusieurs mesures de performances antérieures; la comparaison des performances obtenues (par tel enseignant ou tel établissement ou tel système éducatif) par rapport à l'ensemble des performances. Ce sont là certainement des avancées méthodologiques importantes, curieusement encore peu connues de trop de chercheurs en éducation. Si les mesures de perte ou de gain bruts sont régulièrement utilisées, les mesures de gain ou de perte relatifs (elles relativisent la perte ou le gain bruts en les rapportant à la perte ou au gain maximaux) le sont plus rarement. Plus rarement encore sont employés des indices de rendement multidimensionnel:

toujours sur la base d'au moins deux mesures de performance, on peut calculer un indice de rentabilité relative qui se décompose en quatre indices : le degré de maîtrise finale, le degré d'utilité de l'apprentissage entre les deux mesures, le degré de conservation des acquis de départ et le degré d'efficacité de l'apprentissage effectué (De Ketele, 1980). Sans doute aussi par un déficit de formation méthodologique, trop de chercheurs en éducation ignorent les efforts méthodologiques faits ces dernières années pour mesurer le degré d'efficacité et d'équité⁹ de séquences ou d'institutions ou de systèmes d'éducation ou de formation. La base de ces mesures repose sur le calcul de la droite de régression entre les deux mesures sur l'ensemble des sujets ou objets d'études et sur le calcul des résidus à la droite de régression. En effet, un type de séquence ou un établissement scolaire ou un système éducatif est plus efficace qu'un autre s'il fait faire à des élèves un progrès plus grand que le progrès fait par les autres élèves. Ceci peut s'estimer en calculant la moyenne des résidus à la droite de régression des premiers par rapport à la moyenne des résidus des seconds. De même, si on définit l'équité pédagogique comme le fait que la distance qui existe au départ entre les élèves forts et faibles n'augmente pas, et si possible diminue, ou encore par le fait que les faibles de départ devraient bénéficier de progrès au moins équivalents aux plus forts de départ (Bressoux, 1993 ; Sall & De Ketele, 1997), on peut à l'arrivée comparer la moyenne des résidus des élèves faibles au départ à la moyenne des résidus des élèves forts au départ.

À côté de ces avancées cependant, il ne faut pas oublier que beaucoup de problèmes restent non résolus de façon suffisante. Citons quelques exemples. Dans de nombreuses recherches, il est souhaitable de pouvoir comparer les progrès des élèves sur des épreuves équivalentes ; s'il n'est pas difficile de construire des épreuves comparables sur des savoirs ou des savoir-faire élémentaires, il n'en va pas de même pour des épreuves centrées sur des savoir-faire complexes ou des compétences (De Ketele & Gérard, 2004 et à paraître). Si des référentiels et des standards de performance sont progressivement élaborés avec des démarches méthodologiques de plus en plus éprouvées et sont adoptés par des institutions et même des pays de plus en plus nombreux, il reste bien des difficultés à surmonter dans l'opérationnalisation de certains types de critères et de standards. Malgré l'insistance des travaux théoriques sur l'importance de ne pas se contenter de comparer les performances à des normes de groupe mais de privilégier les comparaisons à des standards de performance, les travaux méthodologiques d'opérationnalisation de ces excellentes intentions sont trop peu nombreux¹⁰.

Quelques commentaires pour conclure

La revue *Mesure et évaluation en éducation* a maintenant plus de vingt ans. Pour fêter ses vingt ans, nos collègues canadiens de l'ADMEE¹¹ ont eu l'excellente idée de digitaliser sur un CD-ROM tous les articles parus pendant cette double décennie. Une analyse de l'évolution des titres est particulièrement significative. Dans les premières années, la plupart des articles étaient essentiellement consacrés à des problèmes de mesure. Ensuite, pendant plus de dix ans, on ne trouve plus guère d'articles de ce type, mais des articles «généralistes» sur l'évaluation ou des articles sur des dispositifs d'évaluation (le thème du portfolio devient d'ailleurs un thème vedette récurrent). Ce n'est que ces dernières années qu'un certain équilibre se rétablit entre les différentes démarches méthodologiques des processus évaluatifs : démarches sommatives (la mesure), descriptive (comme le portfolio), herméneutique (recherche de sens à travers la mise en cohérence d'indices multiples).

On retrouve ce même mouvement historique en France. Dans les décennies qui ont suivi la Seconde Guerre mondiale, les spécialistes en évaluation (comme Piéron, 1963 ; de Landsheere, 1974 ; Noizet & Caverni, 1978 ; Bonniol, 1991) se sont centrés sur la mesure et ont popularisé le concept de docimologie. Par la suite, un mouvement, dont Ardoino a été une figure de proue, s'est créé en réaction aux difficultés mises en évidence par les travaux docimologiques et a cru utile de distinguer et d'opposer le contrôle et l'évaluation. Le contrôle aurait pour fonction d'examiner le degré de conformité des résultats observés (les produits) à des résultats normes par des procédures standardisées (la mesure relève du contrôle). L'évaluation aurait pour fonction de donner du sens non seulement aux produits mais aussi aux processus et aux éléments de contexte dans lequel se déroule l'apprentissage ou la formation. Cette distinction rejoint en grande partie la distinction faite par les collègues canadiens de l'ADMEE entre la mesure et l'évaluation.

Personnellement, nous sommes persuadés que le temps de la synthèse est venu. Opposer mesure et évaluation ou contrôle et évaluation ou encore quantitatif et qualitatif sont des attitudes stériles et stérilisantes. Une note, une liste descriptive, un ensemble d'indices, et le reste, n'ont pas de sens en eux-mêmes ; il faut leur en donner. Donner du sens est la fonction du processus de compréhension. Comprendre intervient dans de nombreux processus cognitifs, comme l'avait bien compris Bloom. Donner du sens n'est donc pas spécifique au processus évaluatif et ne peut donc à lui seul le définir, comme semblent l'indiquer certains collègues français. Il faut donc en revenir à la position de Stufflebeam

(1971) qui disait déjà que la fonction première de l'évaluation est de fonder la prise de décision. En éducation, l'évaluation sert à fonder trois grands types majeurs de décision: certifier, réguler et orienter. Pour chacune de ces fonctions, on dispose de trois grands types de démarches: la mesure (démarche sommative), la description et l'herméneutique. À l'intérieur de chaque fonction et compte tenu de la ou des démarches utilisées, il va de soi qu'il est nécessaire de donner du sens à l'information recueillie en vue de fonder la décision à prendre.

Ceci nous conduit à dire que la démarche herméneutique est importante, car elle consiste à établir un ensemble d'indices pertinents (mesures et indices descriptifs) et leur donner du sens pour fonder un certain type de décision. Nous avons donc besoin de travaux méthodologiques sur la démarche herméneutique en relation étroite avec les fonctions visées par l'évaluation. Si, pour nous, l'éduométrie sera toujours subordonnée à une telle herméneutique, il est également vrai qu'une telle herméneutique (donner du sens pour fonder tel type de décision) ne peut se développer sans avoir une excellente maîtrise de l'éduométrie et une excellente maîtrise des indices descriptifs (non seulement pour ce qui est des produits mais aussi des processus) relatifs à tel ou tel type d'apprentissage et de formation.

Un dernier commentaire se basera sur la constatation «empirique» suivante. Dans ma carrière de chercheur, j'ai lu de nombreuses recherches qualifiées par leurs auteurs de «qualitatives». Les seules qui m'ont paru suffisamment rigoureuses (dans le sens où les conclusions n'étaient pas dépendantes de celui qui a fait l'analyse) étaient, la plupart du temps, le fait d'auteurs ayant aussi une excellente formation quantitative. Ceci rejoint d'ailleurs les propos d'une des méthodologues les plus réputées en sciences humaines, Madeleine Grawitz. Dans son ouvrage couronné par l'Académie des sciences morales et politiques (*Méthodes des sciences sociales*), celle-ci disait déjà, en 1976, que l'opposition au quantitatif est absurde mais trouve sa source dans différents facteurs:

- l'inadaptation du quantitatif à certains objets des sciences humaines;
- l'hostilité de ceux qui ne maîtrisent pas l'outillage mathématique;
- les excès de zèle de certains quantitativistes;
- la confusion entre mathématiques et quantification et l'oubli qu'à côté des mathématiques quantitatives, il existe des mathématiques qualitatives;
- l'oubli qu'au départ, la réalité n'offre, le plus souvent, que des données qualitatives, mais qu'il existe des possibilités de les ordonner et parfois de les quantifier (la quantification repose sur une qualification rigoureuse).

Celle-ci explique très bien qu'une bonne quantification (une bonne mesure) repose sur une très bonne qualification. C'est pourquoi, une bonne formation aux méthodologies de la construction d'une mesure implique une bonne formation à la « qualification ». L'avenir de la recherche en évaluation sera fonction de la qualité de la formation des futurs chercheurs dans ce domaine.

NOTES

1. De telles recherches ont été réalisées par des chercheurs du GIRSEF (Groupe de recherche sur les systèmes éducatifs et de formation) et du SPE (Service de pédagogie expérimentale) des universités de Louvain et de Liège. Elles sont consultables sur leurs sites.
2. Même l'autoévaluation est une action sociale située. Contrairement à un préjugé, l'autoévaluation sous-tend toujours plus d'un acteur.
3. En effet, on fait la somme des points obtenus à des questions différentes ou à différents critères. Dans un autre sens aussi, la note reçue par l'élève peut constituer une « sommation » ou un avertissement.
4. Les travaux menés en France par la Commission sur l'avenir de l'école en France et les rapports commandités par le Haut Conseil de l'évaluation de l'école mettent clairement en évidence qu'il s'agit là d'une question cruciale.
5. Encore faudrait-il s'entendre, tant conceptuellement qu'opérationnellement, sur ce que cela signifie, ce qui est loin d'être le cas !
6. La section belge de l'ADMEE a d'ailleurs consacré, en 2003, une journée à ce thème de la judiciarisation au sein du système éducatif. En Belgique, les parents peuvent contester les décisions certificatives prises par l'école et s'adresser à une instance de recours.
7. À l'exception des enfants et des adolescents relevant d'une éducation spécialisée.
8. Ces aspects méthodologiques sont exposés et détaillés dans l'article précité de Jorro, 2002.
9. À ce sujet, on doit beaucoup aux chercheurs de centre de recherche comme l'Institut de recherche en économie de l'éducation de Dijon ou le Service de pédagogie expérimentale de Liège ou le Groupe interfacultaire de recherche sur les systèmes d'éducation et de formation de Louvain-la-Neuve.
10. Dans son célèbre *Handbook of multivariate experimental psychology*, Cattell (1966) se demandait déjà pourquoi les chercheurs en psychologie se contentaient trop exclusivement de standardiser leurs données en références à la facette sujets, alors que deux autres facettes (les variables et les occasions ou situations) sont aussi importantes. Si le « Cube de Cattell » est resté célèbre, on se demande pourquoi il existe si peu de travaux ayant opérationnalisé deux de ses trois facettes. Dans le domaine de l'éducation, Jean Cardinet (1986) tenait le même raisonnement et suscitait tout un ensemble de travaux sur la théorie de la généralisabilité. Avec Daniel Bain, il a coordonné un numéro spécial de la revue *Mesure et évaluation en éducation* comportant quelques recherches s'appuyant sur un tel modèle. Ces recherches sont cependant encore trop rares.

11. Il est intéressant de noter la signification du sigle ADMEE selon le continent. Nos collègues canadiens le traduisaient, jusqu'il y a peu, par « Association pour le développement de la mesure et de l'évaluation en éducation », tandis que les collègues européens le traduisaient par « Association pour le développement des méthodologies de l'évaluation en éducation ». Ajoutons que lors d'une assemblée générale récente, nos collègues canadiens ont adopté la dénomination européenne, se basant notamment sur l'argument qu'en éducation, la mesure est une des démarches méthodologiques possibles (par ailleurs fréquentes dans les pratiques) de l'évaluation.

RÉFÉRENCES

- Allal, L. (1979). Stratégies d'évaluation formative: conception psycho-pédagogique et modèles d'application. In L. Allal, J. Cardinet & P. Perrenoud, *L'évaluation formative dans un enseignement différencié*. Berne: Peter Lang.
- Bonniol, J.-J. (1981). *Déterminants et mécanismes des comportements d'évaluation des épreuves scolaires*: Thèse de doctorat d'État, Bordeaux.
- Bressoux, P. (1993). *Les effets des écoles et de classes sur l'apprentissage de la lecture*. Thèse de doctorat inédite, Université de Bourgogne, Dijon.
- Bressoux, P., & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris: PUF.
- Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi experimental designs for research*. Chicago: Rand Mac Nally.
- Cardinet, J. (1986). *Évaluation scolaire et mesure*. Bruxelles: Éditions De Boeck.
- Cattel, R.B. (1966). *Handbook of multivariate experimental psychology*. Chicago: Rand Mac Nally.
- Cnudde, M., & De Ketele, J.-M. (sous la direction de) (1986). *Évaluation du cours de mathématique de sixième primaire en deuxième rénové*. Louvain-la-Neuve: mémoire inédit.
- Cnudde, M., De Ketele, J.-M., Deland, G., & Delory, C. (1988). De la sixième primaire à la deuxième rénové, que deviennent les acquis en mathématiques? *Bulletin de psychologie scolaire et d'orientation*, 37(1), 23-43.
- Dauvisis, M.-C. (2001). *Outils et méthodes d'évaluation: quand l'ingénierie sollicite ou questionne la recherche*. Actes du colloque international de l'AECSE, Lille.
- De Bruyne, P., Herman, J., & De Schoutheete, M. (1974). *Dynamique de la recherche en sciences sociales*. Paris: PUF.
- D'Hainaut, L. (1975). *Concepts et méthodes de la statistique*. Bruxelles-Paris: Labor-Nathan.
- De Ketele, J.-M. (1980). *Observer pour éduquer*. Berne-Miami: Éditions Peter Lang.
- De Ketele, J.-M. (1993). L'évaluation conjuguée en paradigmes. *Revue française de pédagogie*, 103, 59-80.
- De Ketele, J.-M. (2002). *Synthèse du Colloque du 20^e anniversaire des zones et réseaux d'éducation prioritaire*. Paris: Ministère de l'Éducation nationale.
- De Ketele, J.-M., Delory, C., Furnal, C., Masset, J., & Warnet, A. (1995). À propos d'un essai d'analyse d'erreurs aux épreuves de mathématiques d'un examen cantonal. *Revue d'orientation scolaire et professionnelle*.
- De Ketele, J.-M., & Gérard, F.-M. (2004). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation*, à paraître (voir aussi [http://www.bief.be/enseignement/publication/index_fr.htm]).

- De Ketele, J.-M., & Roegiers, X. (1991, 1993). *Méthodologie du recueil d'informations*. Paris-Bruxelles: Éditions De Boeck.
- De Landsheere, G. (1974). *Évaluation continue et examens. Précis de docimologie*. Bruxelles-Paris: Labor-Nathan.
- Delannay, L., & De Ketele, J.-M. (sous la direction de) (2004). *De la conservation ou de la non-conservation des acquis en mathématiques*. Louvain-la-Neuve: mémoire inédit.
- Delory, C. (1993). *La rénovation de l'enseignement fondamental: ses effets sur les compétences des enfants en fin d'études primaires*. Louvain-la-Neuve: thèse inédite sous la direction du Professeur J.-M. De Ketele.
- Dupriez, V., & Dumay, X. (2005). L'égalité des chances à l'école: analyse d'un effet spécifique de la structure scolaire. *Les cahiers du GIRSEF*, février 2005.
- Figari, G., & Achouche, M. (2001). *L'activité évaluative réinterrogée*. Paris-Bruxelles: Éditions De Boeck.
- Frères, G., & De Ketele, J.-M. (sous la direction de) (1985). *Le conseil de classe: formations des attitudes chez les enseignants et appréciations sélectives des élèves*. Louvain-la-Neuve: mémoire inédit FOPA.
- Jorro, A. (2000). *L'enseignant et l'évaluation. Des gestes évaluatifs en question*. Bruxelles: Éditions De Boeck.
- Jorro, A. (2002). Comment les acteurs se saisissent de l'évaluation? *En Question, 1*. Aix-en-Provence.
- Grawitz, M. (1976). *Méthodes des sciences sociales* (3^e éd.). Paris: Dalloz.
- Hanushek, E.A., & Wöbman, L. (2005). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *NBER Working Paper Series*, Working paper 11124, [<http://www.nber.org/papers/w11124>].
- Marchand, F. (1979). *Évaluation des élèves et conseil de classe*. Paris: Éditeurs Épi.
- Maroy, C. (éd.) (2002). *L'enseignement secondaire et ses enseignants*. Bruxelles: Éditions De Boeck.
- Noizet, G., & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris: PUF.
- Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques*, 280, 47-60.
- Paquay, L. (sous la direction de) (2004). *L'évaluation des enseignants. Tension et enjeux*. Paris: L'Harmattan.
- Piéron, H. (1963). *Examens et docimologie*. Paris: PUF.
- Roegiers, X. (1997). *Analyser une action d'éducation et de formation*. Bruxelles: Éditions De Boeck.
- Scallon, G. (1988). *L'évaluation formative des apprentissages* (tomes 1 et 2). Québec: PUL.
- Sall, N.H., & De Ketele, J.-M. (1997). L'évaluation du rendement des systèmes éducatifs: approches conceptuelle et problématique. *Mesure et évaluation en éducation*, 19(3).
- Stufflebeam, D., et al. (1971). *Educational evaluation and decision making*. Masca: Peacock Publishing. (Traduit par J. Dumas, *L'évaluation au service de la prise de décision*. Ottawa: Éditions NMP, 1980.)
- Vial, M. (1996). L'autoévaluation, entre autocontrôle et autoquestionnement. *En question, Titre 1*, Aix-en-Provence.
- Vial, M. (2001). *Se former pour évaluer. Se donner une problématique et élaborer des concepts*. Bruxelles: Éditions De Boeck.
- Weiss, J. (1984). *La subjectivité blanchie*. Neuchâtel, IRDP: document inédit IRDP/R 84.17.